R E P O R T R E S U M E S

ED 014 131 EA 000 788

STUDIES AND REPORTS--AN EMPIRICAL ANALYSIS OF A SYSTEM OF ACHIEVEMENT GRADING BASED ON THE DISTRIBUTION OF SCHOLASTIC APTITUDE IN A CLASS.

BY- MASON, GEOFFREY P.
BRITISH COLUMBIA EDUCATIONAL RES. COUNC., VICTORIA

REPORT NUMBER BCERC-R-8 PUB DATE NOV 65
EDRS PRICE MF-$0.25 HC-$0.76 17P.

DESCRIPTORS- STATISTICAL ANALYSIS, *ACHIEVEMENT RATING, *GRADING, *ACADEMIC APTITUDE, *TEACHERS, GRADES (SCHOLASTIC), INTELLIGENCE QUOTIENT, ACHIEVEMENT TESTS, GRADE 6, *ELEMENTARY SCHOOL STUDENTS, ACADEMIC ACHIEVEMENT, VICTORIA,

FINDINGS OF THIS STUDY INDICATE THE NECESSITY OF REEXAMINING THE UNMODIFIED USE OF SCHOLASTIC APTITUDE LETTER GRADE DISTRIBUTION AS THE BASIS FOR THE ACHIEVEMENT LETTER GRADE DISTRIBUTION OF A CLASS. THE USE OF THIS SYSTEM IN BRITISH COLUMBIA IS BASED ON THE ASSUMPTION THAT FOR ANY SIZE SAMPLE OF STUDENTS, THERE WILL BE A HIGH CORRELATION OF I.Q. SCORES AND CLASSROOM ACHIEVEMENT GRADES. THE STUDY REVEALED, HOWEVER, THAT THE EXTENT OF DEVIATION BETWEEN THE TWO WAS CONSIDERABLE. THE ARGUMENT THAT THIS PRACTICE GROUPS THE ACHIEVEMENT LETTER GRADES AROUND A MID-POINT APPROPRIATE FOR THE CLASS WAS FOUND TO BE INCORRECT FOR MANY CLASSES IN THE STUDY. IN BOTH THE HIGH AND LOW SCHOLASTIC APTITUDE GROUPS, ABOUT AS MANY CLASSES WERE FOUND TO UNDER-ACHIEVE AS TO OVER-ACHIEVE IN RELATION TO THEIR SCHOLASTIC APTITUDE MEAN. CERTAIN PROCEDURES MODIFYING THIS METHOD ARE RECOMMENDED FOR TEACHER USE IN GRADING ACHIEVEMENT. (HM)

ED014131

BRITISH COLUMBIA EDUCATIONAL RESEARCH COUNCIL

STUDIES AND REPORTS

AN EMPIRICAL ANALYSIS OF A SYSTEM OF ACHIEVEMENT GRADING BASED ON THE DISTRIBUTION OF SCHOLASTIC APTITUDE IN A CLASS

REPORT NO. 8

by GEOFFREY P. MASON

EA 000 788

BRITISH COLUMBIA EDUCATIONAL RESEARCH COUNCIL

STUDIES AND REPORTS

AN EMPIRICAL ANALYSIS OF A SYSTEM OF ACHIEVEMENT GRADING BASED ON THE DISTRIBUTION OF SCHOLASTIC APTITUDE IN A CLASS

REPORT NO. 8

NOVEMBER, 1965

BY GEOFFREY P. MASON
UNIVERSITY OF VICTORIA

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

This study was made possible through financial support from the University of Victoria Faculty Research Fund. The personal assistance of Miss Verna Turner, Department of Tests and Measurements, Greater Victoria School Board who made records available, and of Mr. Gordon Fearn, who did much of the computation, is greatfully acknowledged.

STUDIES AND REPORTS

BY GEOFFREY P. MASON
UNIVERSITY OF VICTORIA

A periodic and difficult task of the classroom teacher is the evaluation of the achievement of each of his pupils. Essentially, grading a student's work may be accomplished in one or both of two ways. The first involves a comparison of the student's achievement with an absolute standard. This results in a mark or score which may be expressed, for example, as 20 out of 25 words correct in spelling, or 80 percent in arithmetic. The second uses a comparison of the student's achievement with the achievement of other students taking the same work. This can be expressed as a rank in the group, for example, 14 out of 40, or as a percentile rank (out of 100), or as a letter grade which represents a range of several ranks.

The main difficulty in the use of an absolute standard against which to judge the achievement of pupils lies in the establishment of that standard. This is usually done on the basis of the subjective judgement of the examiner, who "feels" that he has set a reasonable standard to be met. The students or other teachers may think otherwise and consider his examination to be very difficult, too easy or unfair. Very rarely do such arbitrary standards have any external validity other than that which resides in the past experience of the teacher with similar pupils taking a similar course.

The problems inherent in establishing absolute standards of achievement by which to judge students, have led educators in British Columbia, to a wide acceptance of the second system, that of stating the rank of a student in his group. To avoid the impression of a greater accuracy than really exists in the allocation of school marks, several ranks are often placed together and a letter grade assigned to the groups of ranks. The system used in British Columbia consists of seven letter grades, A, B, C+, C, C–, D and E, representing a 5, 20, 15, 20, 15, 20, 5 percent distribution. Thus achievement considered to be in the top 5 percent of the class, for example, the top 2 of 40 students, is designated "A" achievement. Achievement in the next 20 percent for example, from the third to the tenth place in a class of 40 students considered to be "B" achievement, that in next 15 percent "C+" achievement, and so on.

While the method of rank ordering students and assigning letter grades to the ranks escapes the problems inherent in imposing arbitrary standards, it nevertheless creates other difficulties in British Columbia. This province has a highly centralized educational system which aims at uniformity of standards throughout. Examinations on a province-wide basis are sometimes set to ensure this uniformity. As a result, it is considered desirable that a given letter grade for achievement in one school district should represent a similar standard in all school districts. In other words, any "A" student or "D" student should be an "A" student or a "D" student in any school of the province, or in any of several classes of a given grade in a given school.

It is apparent that if letter grades are intended to denote the same levels of achievement in any class in the province, the letter grade system previously described cannot be applied without modification. Obviously the top five percent of a homogenously grouped class of good students will differ in achievement from the top five percent of a class of poor students. If a letter grade of A is given to the top five percent of the students in both groups, then the letter grades for each are in no way comparable.

The teacher's problem is now apparent. If he applies the ranking system rigidly, he has no assurance that his letter grades will represent comparable standards to those of other teachers in other classes. If on the other hand he attempts to assess the quality

of one student's work apart from the achievement of other members of the class, he is forced into using the method of an absolute standard with the difficulties previously noted.

In large schools with several classes in one grade, the problem is often met by giving the same examinations to all the classes of a given grade. It is assumed that because of the larger size of the group, the average level of distribution of achievement will be comparable to the overall British Columbia pattern. Two criticisms of this practice may be offered. First, the assumption of average standards and normally distributed achievement in all large schools is incorrect, as some schools have demonstrably better achievement records than others. Second, if the classes in a single grade of a school have been grouped for educational reasons, a common examination tends to vitiate the purposes underlying the grouping and may be considered educationally undesirable. However, irrespective of the validity of these criticisms, the practice of grade wide examinations offers no help to the majority of teachers who teach in schools with no more than one or two classes in each grade. For the majority of classroom teachers there exists the problem of trying to determine how many of their students should justifiably receive a given letter grade from the results of classroom tests.

In an attempt to assist teachers with this problem, Conway (1949) suggested that the distribution of I.Q. letter grades in a class be used as a basis for determining the number of each of the letter grades for achievement. The rationale and procedure as stated by Conway follows:

"....If we make two assumptions, (a) that the influence of the differences between teachers is not very great, and (b) that, on the average, groups of pupils whose scholastic aptitude is high will have high achievement and groups whose intelligence is low will have low achievement, we can devise a method of determining the number of letter grades to assign to any school or class. It may be applied to any group irrespective of its size.

The first step is to convert the pupil's I.Q.'s or M.A.'s to letter grades. This may be done directly from the Tables of the B.C. Norms issued with certain scholastic aptitude tests, or it may be done with somewhat less accuracy from general percentile tables.

If, when we have completed the conversion to letter grades we find that we have 3 A's, 2 B's, 3 C+'s, 4 C's, 6 C–'s, 11 D's and 3 E's according to scholastic aptitude test results, we may safely assume that on the average we also may expect 3 A's, 2 B's, 3 C+'s, etc. in achievement. Therefore when we have given an achievement test we have only to arrange the papers in descending order and to call the top 3 scores 'A', the next 7 'B', etc. This will give an appropriate number of letter grades to each class, irrespective of the difficulty of the examination they have written." (Conway, 1949, p. 2)

In the last few years Conway's proposal has received wide acceptance from educational administrators in British Columbia and the procedure recommended is currently in use in many schools.

Smith (1961) attempted to evaluate the procedure recommended by Conway in an urban district of British Columbia. Smith designated this procedure as the "Standard" method and the awarding of letter grades A, B, C+, C, C–, D and E based on a 5, 20, 15, 20, 15, 20, 5 percent distribution as the "Normal Curve" method. The use of the terms "Standard" method and "Normal Curve" method will be maintained in this paper.

Smith's procedure was to compare two methods of assigning achievement grades as follows:

1. 21 groups were established, each group consisting of all the grade seven students in a given school.

2. Metropolitan Achievement Test results in Arithmetic and Language, and a scholastic aptitude measure from the Pintner General Ability Test were obtained for each student.

3. For each of the 21 groups, achievement letter grades were assigned to each individual's raw score on each of the Metropolitan Achievement sub-tests according to

   a. the "Normal Curve" method.
   b. the "Standard" method.

4. The "Normal Curve" Achievement grades were obtained by taking each of the 21 groups in turn. The individual scores within a group were placed in rank order and the 5, 20, 15, 20, 15, 20, 5 percent divisions made to give the number in each letter grade category.

5. The "Standard" letter grades were obtained according to the procedure suggested by Conway. Local norms expressed in letter grades for the

Pintner General Ability Test were used to determine the number of each letter grade for scholastic aptitude in each group. The same number of each letter grade was then given to the achievement test results of the group.

6. A criterion against which to evaluate these two methods of determining the letter grade was established by converting each achievement test raw score into a letter grade by applying the "Normal Curve" system to the percentile norms issued by the Department of Education of the Province of British Columbia. Thus each achievement score was treated in three ways to convert it to a letter grade. First, by ranking the score in the student's school group and applying the "Normal Curve" percentages. Second, by ranking the score in the school group and applying the number of each letter grade obtained from the scholastic aptitude test. Third, by ranking the score in the results of the whole province and applying the "Normal Curve" percentages to obtain the letter grade. The third method was considered to give the best estimate of the quality of the achievement and was therefore used as the criterion.

7. Finally the agreement of the letter grades in each group obtained by (a) the "Normal Curve" method (b) the "Standard" method with the letter grades obtained by use of Provincial norms was determined.

8. "Agreement" was determined by tabulating the frequency with which the letter grade obtained by each of the two methods was the same as that obtained from the Provincial norms.

Smith reports agreements of 78.8% & 72.3% for language and arithmetic respectively using the "Standard" method compared with 55.6% and 64.3% using the "Normal Curve" method and cautiously suggests that the "Standard" method may be "of value".

However, the groups used by Smith tended to be large. Of the 21 groups used in his study nine were over 100, and seven were nearly 300 or more. Only six had less than 45 students. If the data from these six groups are considered separately the percentage of agreement now becomes 44.7 and 50.0 for language and arithmetic respectively using the "Standard" method, and 68.2 and 51.6 using the "Normal Curve" method, (extracted from Smith, p. 12b.) In other words for these six classes only, the small difference was in favour of the "Normal Curve" method.

Smith's findings unfortunately contribute little to the problem of whether the "Standard" method suggested by Conway can be of value to a classroom teacher in determining grades for his class. Few groups of classroom size were used in his study, and these supported the "Normal Curve" method.

As the value of the "Standard" method is still in question, the present study was undertaken to re-examine Conway's proposal. Part I examines the original statement and the rationale which appears to underlie it, and illustrates the accuracy of prediction by regression estimation of intervals. Part II examines some of the empirical results of the application of the "Standard" method to a large number of grade six classes.

PART I

This section attempts to examine the rationale underlying Conway's proposal to use the scholarship aptitude letter grade distribution as a basis for the achievement letter grade distribution.

Let us assume that there exists a population of 1,000 students to which there have been administered an achievement test and a scholastic aptitude test. The "Normal Curve" percentages have been applied to both sets of results and letter grades awarded to the raw scores. It is readily apparent that, as the same fixed percentages have been applied to the 1,000 children, a correspondingly identical number of each letter grade will have been given for achievement and scholastic aptitude. In other words, 50 students will have received an A grade for achievement and 50 an A grade for scholastic aptitude, 200 a B grade for achievement and 200 a B grade for scholastic aptitude, and so on. Some of the students who have received an A for scholastic aptitude would also have gained an A in achievement, but others would not. Similarly, of the students who have received a B for scholastic aptitude, some might have obtained an A for achievement, others B, and still others C+ or lower. Whatever the extent of agreement between the grades for the two tests for individual children, the number of students in each letter grade category would remain

the same, because these numbers represent given percentages of the student population of 1,000 being considered.

Suppose a sample of forty children from the population of 1,000 is now taken and placed in one class in a school. Will the same relationship between the numbers in each letter grade category on the two tests still obtain? Would one, knowing the numbers of each letter grade awarded on one test, be able to predict, even approximately, the letter grade distribution in the class for the other test? This is the essence of the problem.

Logically, it appears difficult to justify a proposal that, for example, because two students have A scholastic aptitude grades, then two other students in the class be given A grades for achievement. Any justification appears to be contained in Conway's statement that "on the average, groups of pupils whose scholastic aptitude is high will have high achievement and groups whose intelligence is low will have low achievement." (Conway, p. 2), but how such a statement would apply to a class consisting of children of varying ability is not clear.

Both Conway and Smith suggest that the letter grade distribution for scholastic aptitude as the basis for establishing the achievement grades may be modified to accommodate gaps in the distribution of marks, but the extent of the modifications in the examples given by them is very slight. (Conway, p. 4; Smith, p. 9). This does, however raise the question of how much deviation from the scholastic aptitude letter-grade pattern might legitimately be made, and the basis for such a deviation. Assuming that the correlation between scholastic aptitude and achievement is of the order of .60 to .70, what degree of correspondence can be expected between the two distributions of letter grades if they are obtained independently?

For individual students, the problem just posed could be approached theoretically through the use of regression estimation of intervals, which would determine the range of grades for achievement which could be predicted for a given grade of scholastic aptitude. For example, since a B letter grade represents a percentile rank of from 75 to 94, by using a rank order correlation of the magnitude found to exist between scholastic aptitude and academic achievement, an estimate can be made of the percentile interval within which the achievement will fall 95 times in 100 for a 75 percentile scholastic aptitude score. Similarly, the percentile range for achievement can be calculated for scholastic aptitude at the 94 percentile. A combination of the achievement estimates for the 75 and 94 percentiles for scholastic aptitude will give the percentile range for achievement to be expected from a B letter grade for scholastic aptitude at the .05 level of confidence.

In order to obtain for the regression estimates an approximate measure of the relationship existing between scores on a test of scholastic aptitude and scores on tests of academic achievement the following procedure was used. Five classes, totalling 180 students, were selected randomly from the 103 classes used in Part II of this study. Correlations were calculated between the percentile rank obtained from the Otis Self-Administering Test of Mental Ability and the percentile rank obtained from standardized tests of reading, language, arithmetic computation and science. The mean correlation of .66, obtained by using Fisher's Z transformation, was considered to be a reasonable approximation, to use for discussion purposes, of the relationship between a scholastic aptitude and academic achievement.

However, as a correlation coefficient of the order of .66 indicates a large amount of scatter around the regression line, it is to be expected that the overlapping of the predictions for achievement from different letter grades for scholastic aptitude will be very considerable. Table I shows the achievement interval estimates for different grades of scholastic aptitude at the .05 level of confidence, based on a correlation of .66 between scholastic aptitude and achievement.

A reading of Table 8 shows that theoretically, on the basis of a correlation coefficient of .66, a considerable range of grades for achievement should be anticipated for a given scholastic aptitude grade. For example, from Table I it can be seen that for a scholastic aptitude letter grade of C+, which represents a percentile range of 74 to 60, the estimate is that 95 times in 100 the achievement letter grade will be between the percentiles 86 and 36, to give a range of letter grades from B to C–. It should be noted that five times in every hundred, or twice in a class of forty children, the achievement may be expected to fall outside the estimate.

The wide range of letter grades for achievement obtained by regression estimation for a given scholastic-aptitude letter grade clearly indicates the folly of attempting to predict an individual's achievement grades on the basis of his scholastic

aptitude standing alone. It also seems to raise serious doubts concerning the wisdom of using the scholastic aptitude letter-grade distribution as the basis for the achievement letter-grade distribution. For the two distributions to correspond precisely, every deviation in one direction from complete agreement of the scholastic-aptitude and achievement-letter grades would require compensatory deviation in the other direction. In other words, all examples of "over-achievement" by students on the basis of the scholastic aptitude letter grade would require a similar amount of "under-achievement" by other students to occur to compensate for it. Furthermore such "over-achievement" and "under-achievement" would have to occur with certain students only, in order to maintain the balance within the system. For example, if two students with 'A' for scholastic aptitude 'under-achieve' and obtain 'C+' achievement grades, then to maintain the balance two students with 'C+' scholastic aptitude must achieve 'A', or two students with 'B' scholastic aptitude achieve 'A' and two students with 'C+' scholastic aptitude achieve 'B'. The extent to which such adjustments would occur in a group of, say, forty children in one class cannot be estimated, although a priori, such compensation would appear to be a somewhat unlikely occurrence, especially as many of the relevant variables such as the motivation of the children, reading ability, and the effectiveness of previous teaching are likely to be non-randomly distributed from one class to another.

On purely logical grounds there appears to be little support for the proposals of Conway and Smith to use the scholastic aptitude letter grade distribution as the basis for the achievement letter grade distribution. Any merit in their proposal would appear to reside in the extent to which the deviations from complete agreement between the scholastic aptitude and achievement letter grades for individuals would average out within a typical class. This cannot be ascertained theoretically. Fortunately, however, there remains a simple empirical approach to this problem, that of examining school records to see what actually occurs in Practice. This was done in Part II of this study.

PART II

Conway's proposal is based on the assumption that a close correspondence exists between the distribution of letter grades for scholastic aptitude and the distribution of letter grades for achievement in a class. This part of the study is concerned with an examination of the correctness of this assumption through a scrutiny of school records to determine empirically the extent of the discrepancies between the distributions.

There are two requirements for such an approach to the problem of the validity of using the "Standard" method. First, marks for achievement must not be contaminated by teachers' expectations of performance. In other words, the achievement test must either be objective or scored without knowledge of the scholastic aptitude grades. Second, the letter grades derived from the scores on the tests of scholastic aptitude and achievement must be strictly comparable, that is, they must be based on norms derived from the same population. Both these re-

TABLE 1

Predicted Grade Intervals for Achievement for
Different Scholastic Aptitude Letter Grades (r = .66)

Scholastic Aptitude		Estimate for Achievement	
Letter Grade	Range in Percentiles	Range in Percentiles	Letter Grade
A	95 or over	99 – 59	A to C
B	94 – 75	99 – 46	A to C
C+	74 – 60	86 – 36	B to C –
C	59 – 40	76 – 23	B to D
C –	39 – 25	63 – 12	C+to D
D	24 – 5	53 – 1	C to E
E	4 and under	40 – 1	C to E

quirements were met in this study by using the results from a battery of standardized tests with norms developed from the test scores of all the children of a given grade in a given area.

The purpose of this part of the study, then, was to examine the empirical support for using the scholastic aptitude letter grade distribution as a basis for determining achievement grades. This was done by examining the extent of the difference between the letter grade distribution for scholastic aptitude and the letter grade distribution obtained in each of six academic subjects, for each of a large number of classes in one school district. The letter grades in each case were obtained from district norms based on the results of standardized tests of scholastic aptitude and academic achievement.

It is sometimes suggested that, while the scholastic aptitude letter grade distribution may not coincide exactly with the distribution of achievement letter grades, it at least groups the grades around a mid-point which is appropriate for the class. In other words, it is hypothesized that the mean of the scholastic ability letter grades and the mean of the achievement letter grades for a class will correspond closely, even though the two distributions of letter grades may be somewhat different. This hypothesis also was examined.

PROCEDURE

The procedure used in this study is first outlined below and then discussed.

In one class, scores from a test of scholastic aptitude and from tests of achievement in six academic subjects were converted to letter grades using norms established for the school district.

For this class the distribution of the letter grades for scholastic aptitude was compared with the distribution of the letter grades for achievement in one of the academic subjects.

The difference between the number of "A" letter grades obtained by the class on the test of scholastic aptitude and the number of "A" letter grades obtained on the test of achievement was calculated. This was repeated for each letter grade.

This procedure was repeated in the class for a total of six tests of achievement.

The whole procedure was repeated for a total of 103 classes, and the results tabulated.

The mean value of the letter grades for scholastic aptitude was also calculated for each class. This mean value was compared with the mean value of the letter grades for a test of achievement.

Differences between the mean values were calculated and tabulated. This procedure was repeated for a total of four tests of academic achievement.

The above outline delineates the essence of the procedure used. However, certain assumptions and checks were made during the study. A description of the procedure therefore now follows in detail.

The records of all the grade six classes in School District 61 in British Columbia for the years 1961-62 and 1962-63 were analysed. Grade six was chosen because of its middle position in the school grades and because complete results of a test of scholastic aptitude and standardized tests of achievement in language, spelling, reading, arithmetic reasoning, arithmetic computation and science were available.[1] Table 2 shows the size and number of the classes in grade six for the two years combined.

TABLE 2

Size of the 103 Grade Six Classes Used

Class Size	Frequency
43 – 44	3
41 – 42	5
39 – 40	15
37 – 38	12
35 – 36	13
33 – 34	18
31 – 32	11
29 – 30	7
21 – 28	7
14 – 20	12

N = 103

The scores for the test of scholastic aptitude and for each of the tests of achievement were translated into the letter grades, A, B, C+, C, C–, D or E using the district norms. A comparison between the letter grade distribution for scholastic aptitude and the letter grade distribution for a test of achieve-

1. *See Appendix A.*

ment was now possible as both sets of letter grades were based on the population of all grade six children of the one school district.

In order that a comparison between different distributions might be effected, the concept of a deviation was introduced. A deviation is here defined as the absolute difference between the number of a given letter grade obtained by a class on a test of scholastic aptitude and the number of that letter grade obtained by the class on a test of achievement. For example, if three A's are obtained on the test of scholastic aptitude and five A's are obtained on a test of achievement, there is a deviation of two. Thus, if the distribution of letter grades for scholastic aptitude for a class was four A's, nine B's, six C+'s, eight C's, five C–'s, seven D's and one E, and the letter grades from a test of arithmetic computation were three A's, eleven B's, eight C+'s, eight C's, seven C–'s, and three D's, the deviations would be one, two, two, zero, two, four and one respectively, to make a total of 12. The mean of the deviations for the letter grades is the total of the deviations divided by the number of letter grades, in this case, twelve divided by seven, or 1.71. In this example, it could be said that the number of each letter grade obtained for arithmetic computation deviated from the number of each letter grade obtained for scholastic aptitude to the extent of a mean of 1.71 for each letter grade.

Not all of the classes used in this study approached a normal distribution of their scholastic aptitude letter grades. Considerable skewness was apparent, in many of the distributions, due to some extent to homogenous grouping practices in the schools, or the socio-economic basis of the areas from which the students in the different classes were obtained. There was a possibility that the amount by which the grades for scholastic aptitude in a class departed from a normal distribution would have an effect on the extent of the difference between the distribution of the letter grades for scholastic aptitude and the distribution of the letter grades for achievement. Consequently, a test was made as follows:

1. A chi-square test of goodness of fit of the obtained scholastic aptitude letter grade distribution with the theoretical distribution based on the fixed percentages of 5, 20, 15, 20, 15, 20, 5, was made for each of the 103 classes.
2. The classes were then rank ordered by the size of the departure from the theoretical distribution as indicated by the chi-square test.
3. Three groups were formed from the bottom 27, middle 50, and top 26 classes, to represent the low, middle and high deviations from the theoretical distribution, and designated Low, Middle and High Groups respectively.
4. Deviations, as described previously, were calculated to obtain the mean deviation of the letter grades obtained on each of the achievement tests from the letter grades obtained for scholastic aptitude.
5. The mean deviation for all the achievement tests combined was then found for the classes divided into the Low, Middle and High Groups. Table 3 gives the mean deviation of the letter grades in all six of the academic subjects combined for the 103 classes divided into the Low, Middle and High Groups.

TABLE 3

Mean Deviation Per Letter Grade in Six Academic Subjects Combined for the 103 Classes Divided into Low, Middle, and High Groups

	Number of Classes	Mean of Deviations
Low Group	27	2.85
Middle Group	50	2.92
High Group	26	3.09

Table 3 shows an increasing mean deviation per letter grade between the scholastic aptitude letter grades and the achievement letter grades as the scholastic aptitude letter grade distribution departs increasingly from the theoretical distribution. However, the increase from 2.85 to 2.92 to 3.09 is relatively small and was not considered significant for the purposes of this study. Consequently, for the remainder of the analysis, the 103 classes were treated as one group.

One further consideration at this stage was the effect of the different degrees of relationship between scholastic aptitude and achievement in the different academic subjects on the size of the deviations. Consequently, the mean deviation per letter grade was calculated for each academic subject in turn using the 103 classes as one group. Table 4 shows the mean deviation per letter grade in each of the six academic subjects.

It may seem from Table 4 that the mean deviation for the six academic subjects combined was 2.95. The lowest mean deviation in any subject was Reading 2.68, and the highest Science 3.14. These represent departures from the total mean of 2.95 of only approximately eight and seven percent respectively. Consequently, the differences among the six academic subjects were not considered to be of sufficient magnitude to warrant a separate analysis for each academic subject. The results from the six subjects, therefore, were pooled and an overall mean of the deviations in all subjects used to best represent the deviation between the letter grades for scholastic aptitude and academic achievement.

TABLE 4

Mean Deviation Per Letter Grade in
Six Academic Subjects in 103 Classes

Subject	Mean Deviation Per Letter Grade
Language	2.96
Spelling	2.91
Reading	2.68
Arithmetic Reasoning	2.91
Arithmetic Computation	3.10
Science	3.14
Mean of the 6 subjects combined	2.95

Having pooled all classes and all of the six academic subjects, data were collected and tabulated in the following manner. First, all those cases where, in a single class, the letter grade frequency for scholastic aptitude was zero were examined to see how many of the letter grade were actually obtained by the class on the achievement tests. This was repeated for each of the 103 classes. The mean of the deviations from zero was then calculated. The procedure was then repeated for all cases where the letter grade frequency for scholastic aptitude was one, and the mean of the deviations of the achievement frequencies from one was calculated. This was repeated for letter grades for scholastic aptitude with a frequency of two, then for those with a frequency of three, and so on to a frequency of fifteen.

Table 5 gives the mean of all the deviations, the interquartile range of the deviations, and the full range of the deviations for each frequency of letter grade for scholastic aptitude.

The interpretation of Table 5 may be illustrated by taking a specific instance from the table. Suppose the frequency of a given letter grade for a class of scholastic aptitude was 5. Reading down the first column to "5" one finds, that for the 103 classes in this study, where the scholastic aptitude letter grade frequency was 5, the frequencies of the letter grade actually obtained for achievement deviated from 5 to the extent of a mean of 2.32. The middle 50 percent of the deviations were from 1.5 to 3.8 from the frequency of 5, while the more extreme examples ranged from perfect agreement with 5 to a deviation of 12 from the frequency of 5. In other words, in one extreme case where 5 of the letter grade might be "anticipated" from the scholastic aptitude frequency, 17 of the letter grade were actually achieved on a standardized achievement test.

It may be seen from Table 5 that the mean of the deviations increased as the frequency of the scholastic aptitude letter grade increased. This is a result of the greater freedom to vary as the frequency of the letter grade for scholastic aptitude increased. When the frequency of the letter grade for mental ability is zero, deviations in one direction only are possible, whereas for all frequencies greater than zero, deviations can be in two directions.

In order to test the claim that the means of the scholastic aptitude letter grades and achievement letter grades will correspond, the mean of the scholastic aptitude letter grades and the means of the letter grades for each of language, reading, arithmetic computation and science were calculated for all the students combined and for each of the 193 classes separately. For the following calculations, values were assigned to each letter grade as follows:

A = 3.92, B = 3.00, C+ = 2.41, C = 1.96, C– = 1.51, D = 0.92, E = 0.[1]

The means of the letter grades for all students in the 103 classes combined were found to be as follows: scholastic aptitude 1.96, language 1.98, arithmetic computation 1.94, reading 1.95, science 1.96. This provided a check on the norms used to determine the letter grades and showed them to be consistent.

1. *These figures represent the accumulated Z value from the normal curve of the midpoint of each letter grade interval with the zero point set at E.*

TABLE 5

Mean Deviations, Interquartile Range of the Deviations, and Full Range of the Deviations of the Letter Grade Frequencies for Achievement from the Letter Grade Frequencies for Scholastic Aptitude

Frequency According to Scholastic Aptitude	Mean Deviation	Interquartile Range of Deviations	Full Range of Deviations
0	0.74	0.4 – 1.6	0 – 11
1	1.61	1.0 – 2.5	0 – 12
2	1.89	1.2 – 2.9	0 – 12
3	2.08	1.3 – 3.5	0 – 9
4	2.27	1.4 – 3.8	0 – 10
5	2.32	1.5 – 3.8	0 – 12
6	2.44	1.5 – 4.0	0 – 10
7	2.54	1.6 – 4.2	0 – 10
8	2.59	1.3 – 4.0	0 – 8
9	2.71	1.5 – 4.5	0 – 10
10	3.50	2.0 – 5.5	0 – 11
11	3.81	2.0 – 6.4	0 – 8
12	3.94	2.4 – 6.0	0 – 11
13	4.44	2.8 – 7.0	0 – 10
14	4.68	2.6 – 7.4	0 – 12
15	5.48	3.2 – 8.6	0 – 13

TABLE 6

The Mean of the Absolute Deviations of the Mean Achievement Letter Grade from the Mean Scholastic Aptitude Letter Grade for 103 Classes in 4 Academic Subjects

Subject	Mean Deviation per Class
Language	.23
Reading	.15
Arithmetic Computation	.25
Science	.28
Four subjects combined	.23

Each of the 103 classes was then considered separately. An absolute deviation was calculated for each class between the mean of the scholastic aptitude letter grades and the mean of the letter grades for language. Thus, if the mean scholastic aptitude letter grade was 1.80 and the mean language letter grade was 2.08, the deviation was .28 for the class. The mean of the absolute deviations for language was then calculated for the 103 classes. This procedure was repeated for the reading, arithmetic computation, and science letter grades.

Table 6 shows the mean amount by which the mean of the achievement letter grades deviates from the mean of the scholastic aptitude letter grades in each of the 103 classes.

The mean of .23 represents an average shift, for example, for classes with a mean of C scholastic aptitude, of approximately half the distance to a C+ or C– in achievement. With classes with a C+ mean for scholastic aptitude the shift for achievement on the average is nearly halfway to a B or to a C.

In order to illustrate the variability that exists in achievement from one class and subject to another, the results of ten typical classes selected to represent the range of scholastic aptitude are shown in Table 7.

As there are differing degrees of relationship between scholastic aptitude and the various academic subjects an analysis of the size of the class deviations in each of the four academic subjects was

TABLE 7

Means* of the Letter Grades for Scholastic Aptitude, Language, Reading, Arithmetic Computation and Science for 10 Classes

Class	Scholastic Aptitude	Language	Reading	Arithmetic Computation	Science
1.	2.61	2.97	2.81	3.07	2.92
2.	2.57	2.04	2.35	2.11	1.76
3.	2.44	2.76	2.47	1.67	1.27
4.	2.19	2.22	2.04	1.46	2.19
5.	1.82	1.59	1.81	1.07	1.97
6.	1.74	1.81	1.60	1.46	1.57
7.	1.57	2.10	1.59	1.50	1.17
8.	1.32	1.31	1.30	1.03	1.55
9.	1.21	1.30	1.31	1.11	1.28
10.	1.09	1.30	1.27	1.21	.85

* For the calculations the following values were used:
A = 3.92 B = 3.00 C+ = 2.41 C = 1.96 C- = 1.51 D = .92 E = 0

made. Table 8 shows for each academic subject the number of classes which obtained a deviation of a certain size. For example, reading from the table, in language 23 classes out of 103 obtained deviations of from .11 to 20.

Table 8 shows much closer agreement between the means for scholastic aptitude and reading than for the other subjects. For example, only eight of the 103 classes had deviations of their means greater than .30 for reading, whereas, the corresponding figures for language, arithmetic computation and science are 27, 26 and 32 respectively. Nine classes in science had deviations of their means of over .70 which is considerably in excess of one full letter grade.

In view of the considerable variation in deviation from class to class one further question was examined. Does the average level of scholastic aptitude for a class effect the size of the deviation of its achievement mean from its scholastic aptitude

TABLE 8

Frequency of Various Deviations from the Scholastic Aptitude Letter Grade Mean of the Letter Grade Means in Language, Reading, Arithmetic Computation and Science for 103 Classes

Size of Deviation	Language	Reading	Arithmetic Computation	Science
0 – .10	33	45	23	16
.11 – .20	23	31	31	33
.21 – .30	20	19	23	22
.31 – .40	4	6	4	13
.41 – .50	4		12	6
.51 – .60	17		2	2
.61 – .70	2	2	4	2
.71 – and over	0		4	9
	103	103	103	103

For the calculations the following values were used:
A = 3.92 B = 3.00 C+ = 2.41 C = 1.96 C- = 1.51 D = .92 E = 0

mean? In other words, do brighter classes more nearly fulfill expectations on the basis of scholastic aptitude than slower groups, or does the reverse obtain?

To answer this question the 103 classes were divided into three groups on the basis of the mean of the scholastic ability letter grades for the class. The 34 classes with scholastic aptitude means above 2.18 formed the High Group, the 34 classes with means below 1.75 formed the Low Group, and the remaining 35 classes the Middle Group. The mean deviations in each of the four subjects were computed for each group. The results are shown in Table 9.

The average size of the deviations in each of the three groups appears to answer the question previously raised. For both language and reading the deviations are similar for each of the groups. However, for arithmetic computation and science there are considerable differences between the High and Low Groups. In other words the Low Group conforms more closely to expectation on the basis of scholastic aptitude than does the High Group, although the average amount of deviation for the Low Group in Science is still nearly half a letter grade.

The raw data were inspected to see whether there was any consistency in the direction of the deviations for the High and Low Groups. The directions of the deviations are as follows: High Group, arithmetic computation 19 plus, 15 minus; science 19 plus, 15 minus; Low Group, arithmetic computation 14 plus, 20 minus; science 22 plus, 12 minus. Thus it may be seen that in both the High and the Low groups for the two subjects combined roughly half the classes over-achieved and half under-achieved on the basis of the scholastic aptitude mean.

DISCUSSION AND RECOMMENDATIONS

In the light of the empirical findings shown in Table 5 any close adherence to a system whereby achievement letter grade distributions are based on scholastic aptitude letter grade distributions would be obviously inappropriate for the vast majority of the 103 classes studied. The interquartile range of the deviations shows that considerable deviations existed even for the middle 50 percent of the classes, which certainly cannot be considered atypical. In more unusual cases, that is, the 25 percent of the classes above the interquartile range, deviations were often very large. From the last column of Table 5, which gives the full range of the deviations, the magnitude of the more extreme cases can be inferred. For example, when the frequency of a letter grade for scholastic aptitude in a class was seven, the deviations ran from zero, represented by a frequency of seven, to ten, represented by a frequency of seventeen. When the frequency of the letter grade for scholastic aptitude was 14, deviations ran from zero, i.e. perfect agreement, to 12, represented by a frequency of the achieve-letter grade of either 2 or 26. In other words, Table 5 shows one should not expect anything approaching perfect agreement between the number of a scholastic aptitude letter grade in a class and the number of that letter grade for achievement, for any of the frequencies from zero to 15 of the letter grade for scholastic aptitude.

The differences between the scholastic aptitude and achievement letter grade frequencies shown in Table 5 raise serious doubts concerning the validity of using the scholastic aptitude letter grade distribution as the basis for the allocation of achievement letter grades in a class. An examination of the data shows that a close correspondence between

TABLE 9

Mean Deviations of Means for Achievement from Means for Scholastic Aptitude for Classes in High, Medium and Low Scholastic Aptitude Groups in Four Academic Subjects

Scholastic Aptitude		Deviation of Letter Grade Mean			
Class Mean	N	Language	Reading	Arithmetic Computation	Science
2.19 – 3.20	34	.25	.13	.31	.36
1.75 – 2.18	35	.21	.17	.28	.24
.94 – 1.74	34	.23	.16	.17	.23
Total Mean		.23	.15	.25	.28

the scholastic aptitude and the achievement letter letter grade distribution is a relatively rare occurrence. However, experience has shown that many classroom teachers, lacking opportunities for continual comparisons with large numbers of children, need some form of guidance in the allocation of achievement letter grades to a class. Is the "scholastic aptitude basis for achievement" system then better than no guidance at all? It might be advisable to consider if other satisfactory procedures are available.

Logically, one might assume that the distribution of letter grades obtained from standardized tests of achievement would provide a better basis for grading other tests of achievement than would tests of scholastic aptitude. The main difficulty here is the availability of standardized tests in the different subject areas which are valid for the particular children being tested. Even if suitable tests were available, changes in curriculum content, objectives, or the ordering of topics may render large parts of a test invalid at a certain time in the school year. This particular difficulty can be overcome only by school districts instituting programmes entailing constant revision of their standardized tests and the accompanying norms.

The previous year's work is probably the best predictor of achievement for any one school year. However, the same difficulty concerning the lack of objective measurement exists. To use measures of doubtful validity from one year as a means of estimating achievement in the following year is of little help. Only if the earlier year's achievement has been accurately and validly assessed can such measures be of use.

In the absence of suitable standardized tests of achievement, tests of scholastic aptitude provide a readily available objective measure on which to base predictions of achievement. It has been seen that the achievement of a class rarely conforms closely to the scholastic aptitude letter grade pattern, but could modifications be introduced to increase the validity of the procedure of using the scholastic aptitude letter grade distribution as the basis for predicting achievement?

Let us assume that the middle 50 percent of our deviations, which are described as the "interquartile range" in Table 5, come from usual or typical cases in which achievement level is fairly consistent with the scholastic aptitude level. The bottom 25 percent of cases are those more unusual cases where the scholastic aptitude and achievement letter grades conform closely, whereas those cases above the interquartile range are those where a considerable divergence between scholastic aptitude and achievement exists.

If the middle 50 percent or "typical cases" and the bottom 25 percent or closely conforming cases are grouped together, and the figures rounded and tabulated, the results appear as in Table 10. These figures should be of use to a competent teacher with a "typical" class. A "typical class" is defined as one in which there is reason to believe there is no great disparity between the scholastic aptitude and achievement levels. As the frequencies have been derived empirically from a large number of classes they should provide a sound basis for estimating the number of each achievement letter grade to be awarded. For a teacher with a typical class, then, something approaching the figures given in Table 10 is suggested.

TABLE 10

Suggested Number of a Letter Grade to be Awarded for Achievement In a Class For a Given Number of the Letter Grade Obtained for Scholastic Aptitude

Frequency of a Letter Grade for Scholastic Aptitude	Number of the Letter Grade to be Awarded for Achievement in a Typical Class
0	0 to 2
1	0 to 4
2	0 to 5
3 or 4	up to 100% deviation either way
5 to 7	deviation up to 4 either way
8 and above	up to 50% deviation either way

Table 10 should be read as follows: if the scholastic aptitude frequency of any letter grade for the class is zero, then from 0 to 2 of that letter grade may be awarded to a typical class for achievement; if the scholastic aptitude frequency for any letter grade is one, then 0, 1, 2, 3 or 4 of that letter grade may be awarded for achievement; if the scholastic aptitude letter grade frequency is 3, then a deviation of up to 100% of 3 in either direction is acceptable, i.e. from 0 to 6; if the scholastic aptitude letter grade frequency is 8, then a deviation of up of 50% of 8 in either direction is acceptable, i.e. from 4 to 12.

As Table 5 shows, considerable deviations are to be expected for 75 percent of the cases. Only those cases in the bottom quartile have deviations for most frequencies as small as one or two. Hence, a teacher following the suggestions in Table 10 might be expected to deviate from, more often than closely conform to, the scholastic aptitude letter grade frequency in awarding achievement letter grades.

The adoption of the above suggestions require the teacher to exercise some judgment as to where in the distribution of raw scores changes in letter grade should occur. While it must be recognized that raw scores for most classroom tests fall on an ordinal rather than on a interval scale, for a well construced test the size of the gaps in the distribution of raw scores is the best estimate of the significance of changes in the quantity of the attribute being measured. Large gaps in the raw score distribution suggest changes in the underlying attribute and provide indications of where changes from one letter grade to another might legitimately be made. In addition, the teacher should use his own judgment of the quality of the work, partly perhaps, by considering the previous performance of the pupils concerned. It is important that the teacher be left free to exercise some judgment concerning the quality of the work of his pupils, so that errors due to the rigid application of an inexact system may be avoided.

This in turn raises one final problem for consideration. Teachers, if left completely unrestricted in their assigning of grades sometimes tend to be somewhat overgenerous and produce average grades for achievement that are considerably higher than can be justified. The freedom permitted by the recommendations of Table 10 might possibly result in a considerable shift from the mean C letter grade.

The findings with regard to the difference in class means between the scholastic aptitude and achievement letter grades should be of assistance. An inspection of Table 7 and 8 shows that the mean for scholastic aptitude and the mean for achievement often differ substantially. For example, Table 7 shows that Class 3 has a scholastic aptitude mean of 2.44 which is almost exactly C+. The mean of the letter grades for language for this same class is 2.76 or half-way between C+ and B; for reading 2.47 which is approximately C+; for arithmetic computation 1.67 which is slightly above C−; for science 1.27 which is between C− and D. On the other hand the average achievement for class 9 is in substantial agreement with the average scholastic aptitude showing deviations of language .09, reading .10, arithmetic computation .10 and science .07. As Table 6 shows the mean deviation in each of these subjects for the 103 classes was .23, .15, .25, and .28 respectively.

While an average difference between the means for scholastic aptitude and achievement of .23 shown in Table 6 represents a shift of only approximately half a letter grade, 13 percent of the deviations were at least .45 (Table 8), which represents a shift for classes with C average scholastic aptitude of a full letter grade. It should be noted that the mean of the scholastic aptitude letter grades may, once in seven or eight times, overestimate or underestimate the mean of the achievement letter grades to the extent of a full letter grade in the C range. For the 103 classes studied, 61 percent of the deviations were found to be within half a letter grade.

Recognizing on one hand that the classroom teacher often needs guide lines within which to operate, and on the other hand the desirability of leaving the teacher free to exercise professional judgment, a procedure is now suggested on the basis of the findings of this study. This procedure, while based on the original suggestion of Conway, utilizes the findings concerning the mean achievement of a class relative to its scholastic aptitude mean, and the evidence of the considerable variations that occur between the distributions of letter grades for scholastic aptitude and achievement.

The suggested procedure for achievement letter grading within a class is as follows:

1. The classroom teacher must first decide if the mean of the letter grades for achievement for his class should equal the mean of the scholastic aptitude letter grades. If not, he must decide how much deviation should be allowed and in what direction. The performance of the class on recent standardized tests, and the teacher's own judgment of the level of the achievement possibly should both be used. The principal of the school might well require a justification of this decision, bearing in mind the data in Table 8.

2. Once having fixed the mean grade value for the class, the suggestions contained in Table 10 could then be applied using the gaps and tied scores in the distribution of marks.

3. The letter grades for achievement should be adjusted so that their mean is approximately equal to the value set in Step 1. In making the computation the values of each letter grade can be rounded conveniently as follows: A = 4, B = 3, C+ = 2.5, C = 2, C– = 1.5, D = 1, E = 0. The necessity to maintain a given mean will counteract any tendency for too many of the deviations from the scholastic aptitude distribution to be in the same direction without justification.

The advantages of the recommended procedure appear to be threefold:

1. It permits the use of breaks in the mark distribution.

2. It makes allowances for high, middle and low achieving classes.

3. It enables the extent of these allowances to be easily examined and controlled with a few relatively simple calculations.

In conclusion, it should be noted that these findings were obtained from 103 classes in one grade of one school district. Although there appears to be little reason to expect a closer correspondence between scholastic aptitude and achievement grades at other levels, the limited nature of the data must make any proposals tentative.

SUMMARY

The findings of this study indicate the necessity for a re-examination of the practice of using without modification the scholastic aptitude letter grade distribution as the basis for the achievement letter grade distribution of a class. The extent of the deviations of the achievement letter grade distribution from the scholastic aptitude letter grade distribution was found to be very considerable (Table 5). The argument that the practice groups the achievement letter grades around a mid-point that is appropriate for the class was examined and found to be incorrect for many of the classes in the study (Tables 6, 7 and 8). In both the high and low scholastic aptitude groups about as many classes were found to underachieve as to over-achieve in relation to their scholastic aptitude mean.

The problem of permitting teachers unrestricted freedom in the assigning of grades was noted, as well as the desirability of the exercise of professional judgment. Consequently, on the basis of 75 percent of the 103 classes studied, certain procedures modifying the "Standard" procedure suggested by Conway, were recommended for the use of classroom teachers when grading achievement.

APPENDIX A

The tests used in this study were:

1961 – 62

1. Otis Self-Administering Test of Mental Ability, Intermediate, Form C.
2. Greater Victoria Language Test, Form C.
3. Greater Victoria Spelling and Dictation Test, Form C.
4. Gates Basic Reading Test, Level of Comprehension, Form 1.
5. Sequential Tests of Educational Progress, Mathematics, Level 4, Form A.
6. Greater Victoria Arithmetic Computation Test, Form C.
7. Greater Victoria Science Test, Form A.

1962 – 63

1 to 4 as for 1961 – 62.

5. Form B.
6. Form D.
7. Form B.

REFERENCES

1. Conway, C.B., A simplified method of approximating scaled scores. Division of Tests and Standards, Department of Education, British Columbia, Memorandum to District Superintendents, 1949 (Rev. 1961).
2. Smith, Robin N., Standards of Achievement in Local examinations. British Columbia Educational Research Council Studies and Reports No. 4, 196 .